A Unified Algorithm for Adaptive Pattern Classification

A Unified Algorithm for Adaptive Pattern Classification

Y. Kumagai

Abstract

Supervised and nonsupervised algorithms have been developed for adaptive pattern classification, respectively. These algorithms, however, have been studied under somewhat restrictive conditions on the probability structure, and, moreover, have some difficult problems in the applications to practical pattern classification systems.

This paper describes a unified algorithm which combines these two algorithms to remove these problems.

1. Introduction

The problem of pattern recognition has recently obtained great attention and many approaches have been proposed for pattern recognition. In many of their approaches pattern classification may be viewed as a problem in statistical classification. In general, however, we have little or no a priori information about the probability structure of patterns. Consequently, adaptive procedures would be required for learning the unknown probability structure.

Supervised and nonsupervised algorithms have been developed for adaptive pattern classification, respectively. These algorithms, however, have been studied under somewhat restrictive conditions on the probability structure. Furthermore, under general conditions, supervised algorithms would require a large number of training patterns for optimal classification. On the other hand, nonsupervised algorithms have a serious disadvantage concerning the rate of convergence. Therefore, there would have some difficult problems in the applications of these algorithms to practical pattern classification systems.

This paper describes a unified algorithm which combines these two algorithms to remove the above problems.

2. A Unified Algorithm

It is generally assumed that each pattern may be represented by an n-dimensional vector. In this paper, however, for ease of notation we consider each pattern represented by a scalar. We deal with the more general cases where a probability density and an a priori probability associated with each pattern class are all unknown, where no assumption need be made about the forms of the probability densities, and where they may be sufficiently overlapping.

Let there be M possible classes ω_1, ω_2, , ω_M. A probability density $p(x|\omega_i)$ and an a priori probability of occurrence P_{ω_i} are associated with each pattern

class. We assume here that each probability density $p(x|\omega_i)$ can be approximated by a finite series of known orthogonal functions:

$$\hat{p}(x|\omega_i) = \sum_{j=1}^{N_i} a_{ij}\varphi_{ij}(x) \qquad (1)$$

where the $\varphi_{ij}(x)$ are known orthogonal functions and the a_{ij} are unknown coefficients.

In this algorithm, we first make use of a supervised algorithm. The supervised algorithm is as follows:

According to practical situations, an appropriate training period is to be determined. The supervised algorithm will adjust the unknown coefficients a_{ij} using a proper set of training patterns. As a measure of the approximation, we take a quadratic measure. Furthermore, the normalization condition must be setisfied for $\hat{p}(x|\omega_i)$. Consequently, we have the following Lagrangian:

$$L_{1i} = \int_X \left\{p(x|\omega_i) - \hat{p}(x|\omega_i)\right\}^2 dx - \lambda_i \left\{\sum_{j=1}^{N_i} a_{ij}d_{ij} - 1\right\} \qquad \text{for all } i \qquad (2)$$

where λ_i are Lagrange multipliers, $d_{ij} = \int_X \varphi_{ij}(x)dx$.

Therefore, optimal coefficients a_{ij}^* can be obtained by minimizing L_{1i}. Then, we set the gradient of L_{1i}, equal to zero, that is,

$$\nabla L_{1i} = \nabla\left[\int_X \left\{p(x|\omega_i) - \hat{p}(x|\omega_i)\right\}^2 dx - \lambda_i \left\{\sum_{j=1}^{N_i} a_{ij}d_{ij} - 1\right\}\right] = 0 \qquad \text{for all } i \qquad (3)$$

Consequently, by solving Eq. (3) the optimal coefficients a_{ij}^* may be given by

$$a_{ij}^* = E^{(i)}\left\{\varphi_{ij}(x)\right\} + \left\{1 - \sum_{j=1}^{N_i} E^{(i)}\left\{\varphi_{ij}(x)\right\}d_{ij}\right\} d_{ij} \Big/ \sum_{j=1}^{N_i} d_{ij}^2 \qquad \text{for all } i \text{ and } j \qquad (4)$$

where $E^{(i)}\{\ \}$ denotes the average over the population of pattern class i.

Let $a_{ij}(k)$ be the estimates of the unknown coefficients a_{ij} at the k-th step of training process. According to Eq. (4) the reasonable estimates of the a_{ij} at the k-th step may be represented by

$$a_{ij}(k) = \sum_{l=1}^{n_i} \varphi_{ij}(x_l)/n_i + \left[1 - \sum_{j=1}^{N_i}\left\{\sum_{l=1}^{n_i} \varphi_{ij}(x_l)d_{ij}/n_i\right\}\right] d_{ij} \Big/ \sum_{j=1}^{N_i} d_{ij}^2 \qquad \text{for all } i \text{ and } j \qquad (5)$$

where n_i denotes the number of training patterns belonging to pattern class i until the k-th step.

For the a priori probabilities, let $\hat{p}_{\omega_i}(k)$ be the estimate of the a priori probability of pattern class i at the k-th step. Then, the reasonable estimates for the unknown a priori probabilities might be

$$\hat{p}_{\omega_i}(k) = n_i/k \qquad \text{for all } i \qquad (6)$$

After a finite number of adjustments of the coefficients a_{ij} during the pre-

determined training period, it does not necessarily follow that the estimates for the coefficients a_{ij} have converged enough to the values of the optimal coefficients a_{ij}^*. Therefore, the supervised algorithm need be followed by some nonsupervised algorithm to continue the adjustment process.

The nonsupervised algorithm in this unified algorithm is as follows:

By using Eq. (1) the over- all probability density of patterns can be approximated as

$$\begin{aligned}\hat{p}(x) &= \sum_{i=1}^{M} \hat{p}_{\omega_i} \hat{p}(x|\omega_i) \\ &= \sum_{i=1}^{M} \hat{p}_{\omega_i} \sum_{j=1}^{N_i} a_{ij}\varphi_{ij}(x) \end{aligned} \tag{7}$$

where $\hat{p}_{\omega_i}$ denotes the estimate of the a priori probability of pattern class i.

We again take a quadratic measure. Furthermore, we must have the following constraints so that we may identify reasonably each probability density and each a priori probability:

$$\int_X \sum_{j=1}^{N_i} a_{ij}\varphi_{ij}(x)dx = \sum_{j=1}^{N_i} a_{ij}d_{ij} = 1 \qquad \text{for all } i \tag{8}$$

$$\sum_{i=1}^{M} \hat{p}_{\omega_i} = 1 \tag{9}$$

Therefore, we have the following Lagrangian:

$$L_1 = \int_X \left\{p(x)\hat{p}(x)\right\}^2 dx + \sum_{i=1}^{M} \lambda_i \left\{\sum_{j=1}^{N_i} a_{ij}d_{ij} - 1\right\} + \mu\left\{\sum_{i=1}^{M} \hat{p}_{\omega_i} - 1\right\} \tag{10}$$

where λ_i and μ are Lagrange multipliers.

Consequently, in order to obtain the optimal coefficients a_{ij}^* and the optimal estimates $\hat{p}_{\omega_i}^*$ of the a priori probabilities, we set the gradient of L_2 equal to zero, that is,

$$\nabla L_2 = \nabla\left[\int_X \left\{p(x) - \hat{p}(x)\right\}^2 dx + \sum_{i=1}^{M} \lambda_i \left\{\sum_{j=1}^{N_i} a_{ij}d_{ij} - 1\right\} + \mu\left\{\sum_{i=1}^{M} \hat{p}_{\omega_i} - 1\right\}\right] = 0 \tag{11}$$

Evidently, Eq. (11) are not linear equations of the unknown parameters. We make use of gradient method to solve Eq. (11). With the use of gradient method to solve Eq. (11). With the use of digital computer we can obtain comparatively easily the reasonable estimates of the unknown parameters. Furthermore, we can obtain definitely the optimal parameters of each pattern class as the numbes of observed patterns approaches infinity.

In this unified algorithm, maximum-likelihood criterion is used for optimal classification of observed patterns. If, for any observed pattern x

$$\hat{p}_{\omega_i}\hat{p}(x|\omega_i) \geqq \hat{p}_{\omega_j}\hat{p}(x|\omega_j) \qquad \text{for all } j,\ j \neq i \tag{12}$$

the pattern x may be assigned to pattern class i. If in Eq. (12) the sign of

equality holds, it is assigned to the pattern class with the smallest index.

3. Conclusion

We proposed a unified algorithm for adaptive pattern classification. This algorithm works under more general conditions and has some advantages for practical applications. Furthermore, it can follow slow changes in the probability structure of pattern classes.

Acknowledgment

The author is grateful to Prof. S. Kitamura for his encouragement.

(Received May 20, 1971)

References

1) K. Hara and Y. Kumagai: "A unified algorithm for adaptive pattern classification" Hawai International conference on system science, Jan. 1969.
2) J. Zaski: "On the probability density estimation" Proc. I.E.E.E. vol 56, pp 866-867, May. 1968.